



## Towards video quality metrics for HDTV

Stéphane Péchard, Sylvain Tourancheau, Patrick Le Callet, Mathieu Carnec,  
Dominique Barba

### ► To cite this version:

Stéphane Péchard, Sylvain Tourancheau, Patrick Le Callet, Mathieu Carnec, Dominique Barba. Towards video quality metrics for HDTV. Second International Workshop on Video Processing and Quality Metrics for Consumer Electronics, VPQM'06, Jan 2006, Scottsdale, AZ, United States. paper 305. hal-00255978

**HAL Id: hal-00255978**

**<https://hal.science/hal-00255978>**

Submitted on 14 Feb 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# TOWARDS VIDEO QUALITY METRICS FOR HDTV

*Stéphane Péchard, Sylvain Tourancheau, Patrick Le Callet, Mathieu Carnec, Dominique Barba*

Institut de recherche en communication et cybernétique de Nantes (IRCCyN)

Équipe Image et vidéocommunication – UMR CNRS 6597

Polytech’Nantes, rue Christian Pauc, La Chantrerie

BP 60601, 44306 Nantes, Cedex, France

## ABSTRACT

High Definition Television (HDTV) is the new generation broadcasting system, proposing a deeper immersion in action and better visualization comfort. New material is used (flat television screens, fast connection, high media storage, etc.) allowing better user experience than standard definition television (SDTV). Higher hardware prices will only be accepted if noticeable higher quality is reached. However, new features bring specific distortions but also enhancements for global quality. New techniques of quality assessment need to be developed in order to characterize the impact of each element in the HDTV chain. The last step of the quality measurement is the complex combination of every quality contribution providing experience quality measurement for the whole system.

## 1. INTRODUCTION

Television has always suffered from a lack of presence, immediacy and impact compared with “cinema-like” experience. Subjective tests [1, 2] have shown that ideal distance in order to watch moving pictures is about three times the screen’s height (four times for programs with rapid movements). The corresponding viewing angle of 20–30° reduces considerably the sensation of presence of the display system. Furthermore, pictures are perceived deep and natural. However, nearer the screen we are, higher the defects of pictures are perceived, particularly the scanning line structure. Therefore, the basic idea to produce psychological effects is to widen the display screen and, simultaneously, to increase considerably the resolution of the source. According to these considerations, High Definition Television (HDTV) has been developed in USA, Canada, Japan, Korea and Australia in the last ten years and is expected in west Europe and China in 2006.

A HDTV broadcasting system is supposed to raise users expectations to a new level. Such a promising market implies the use of new material with the last broadcasting and compression techniques. As HDTV broadcasters want to

assess the whole experience brought to the customer by the service and specifically reach the best service acceptance, new quality metrics need to be developed. The main problem of HDTV introduction is the lack of reference, because too much features have changed compared to SDTV. Therefore, designing a global quality of experience measure for HDTV is an ambitious and appealing project, which need to be performed progressively.

In this paper, we point out the new requirements for subjective quality assessment of HDTV. Then, we propose some ad hoc quality subjective tests protocols applied here to quality comparison between HDTV and SDTV. The purpose of such tests is to identify the quality gap between the new service and the still performing standard definition television. Finally, we investigate through psychophysics experiments a plausible model for one of the most important new artifact inherent with LCD displays: motion blur.

## 2. WHAT’S NEW WITH HDTV

High Definition displays and pay-per-view services imply an increasing budget for consumers of television. It should mean a noticeable higher quality level for the observer. This leads to the need of very precise visual quality measurements. Therefore, HDTV quality metrics should mainly address high level quality range compared to usual quality metrics. As global quality has to be above any reproach, every artifact appearance will be severely punished by observers. Quality assessment is more critical, reducing quality range to the upper part of the quality scale.

### 2.1. Display technologies

Another problem comes with new display technology. With almost a century of existence, CRT display is a mature technology, optimally adapted to display television programs. But as the screen size was heightening, standard CRT displays became bulky and heavy. Among modern flat panel technologies, LCD (Liquid Crystal Display) is the most often used these days and considered as a mature technology.

But this ten-year-old LCD technology has still to be studied and modified before to be considered as an well fitted TV-display technology. However, its qualities in terms of load, design, size and price should lead it at the top of TV-displays in the coming years. Recently, subjective tests have been conducted in order to compare quality picture between CRT and LCD [3]. Most of the 36 video-expert observers evaluated that picture quality on LCD was lower than on CRT. Therefore LCD still needs some improvements before replacing CRT device in quality field.

Among all the defects detected by expert observers, in particular motion blur is still an annoying artifact for moving pictures with rather quick movement. We will address this specific artifact in section 4. Some other shortcomings have been reported by the viewers. First is that LCD could not express the delicate differences in dark areas: black portions look glossy or lighter than on CRT and even reddish or greenish in certain cases. Mura defects can also appears on LCD monitors, they're particularly annoying on flat portions of images. Differences in reproducing colors have also been observed between CRT and LCD, particularly in flesh colors. Observers also noticed a lack of depth-feel in images displayed on LCD. CRT produces natural impression and textures, while on LCD, images are displayed too sharp leading to unnatural perspective. When asked, observers express their discomfort for all these defects and particularly with the black level and, for specific sequences, with motion blur. A big challenge is to design some post/pre processing in order to enhance image in terms of sharpness or colorfulness. This justifies the needs of specific quality metric to measure performance of these processings.

## 2.2. HDTV Formats

Nowadays, HDTV is technically exploited in two definitions:  $1920 \times 1080$  in interlaced mode and  $1280 \times 720$  in progressive mode. Both use MPEG-2 compression: this is first generation HDTV. Subjective tests [4] tend to prove that 720p (for  $1280 \times 720$  progressive) is visually better, but 1080i is still possible to be used in broadcasting production. Moreover, consumers would likely be influenced by the "the larger image the better" effect and prefer 1080i over 720p.

Second generation of HDTV is at present developed and will use the same definitions but with H.264/MPEG-4 AVC compression in order to decrease the bitrate. HDTV quality metric should then be able to be suited with the H.264 coding scheme. Furthermore, deinterlacing is still a challenging issue for HDTV since interlaced format is still used and imply specific distortions. Therefore tools are needed to assess the visual quality induced by such processings.

Later, third generation would use  $1920 \times 1080$  in progressive mode. It is not available yet because of a lack of capture material and of important investments made on first and second generations.

## 2.3. Distortions combination

New material like flat screens and MPEG-4 coding induce proper distortions and also by combination. At the same time, some elements of the system may include processing dedicated to global quality enhancement. These treatments may soften some distortions and reinforce some others. Interactions between all these distortions and enhancements in terms of quality is a fundamental issue of a HDTV broadcasting service.

Finally, the complexity of this multidimensional quality system has to be integrate in a single global quality measurement. In such a metric, the output sequence may get a higher quality measure than the original one. Effectively, quality has a complex global evolution along the system which may lead it over its original value. HDTV quality metric is then much more complex than usual fidelity metrics.

## 3. SUBJECTIVE PROTOCOL: HDTV VS SDTV

### 3.1. Objective

To measure the impact of high definition on the observation distance and on users quality expectations, we have designed a subjective tests protocol called "HD versus SD". The main goal of this protocol is to determine user preferences between HD and SD contents, everything else remaining the same. Basically, it is to know if to look at a bigger image size is more important for the observer than perceiving some artifacts in it. It is also the occasion to measure the average preferred observation distance for HDTV content. This information is a very important parameter since some distortions may be visible at a certain distance and invisible at another one. The criterion for determination of this ideal distance is observers comfort of visualization.

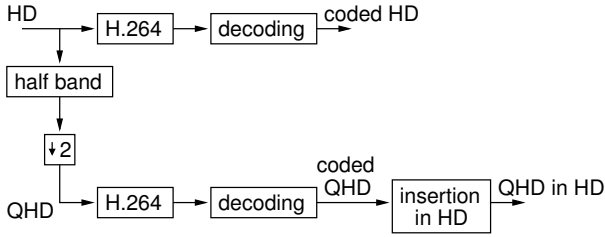
Then is the actual comparison between HD and SD content. This information is important for broadcasters to determine the best HD coding and broadcasting parameters in comparison to what they are doing up to now with SDTV. Here, observers are asked to watch and compare two versions of the same content, one in HD, the other in SD. The SD video is obtained by sub-sampling an HD video and inserting the result in an HD format (see 3.2 for the specifications of the video material). This insertion permits the observer to remain in the same visualization conditions no matter the resolution is SD or HD. Indeed, ITU recommends a viewing distance of three times the height of the image for HD content and six times for SD content.

### 3.2. Video Material

Obtaining good video materials is always a difficult task because of copyrights management, especially in HDTV



**Fig. 1.** Example of HDTV contents usable for “HD vs SD” protocol.

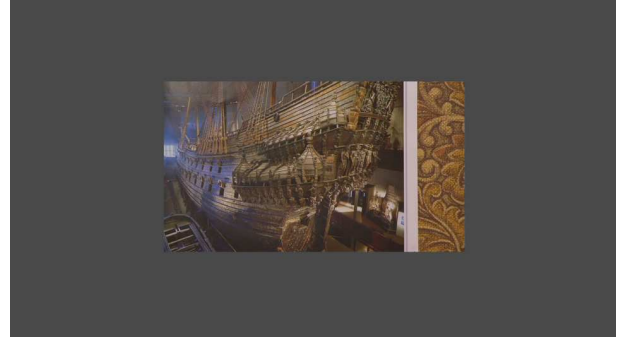


**Fig. 2.** “HD vs SD” protocol contents obtaining scheme.

field. For this tests protocol, we use sequences from SVT researches [4]. Those are four 1080i ( $1920 \times 1080$  interlaced) contents whose first frames are presented in Figure 1. We first distort all these sequences through the use of H.264 JM reference software [5]. Seven bitrates have been used per HD content. Bitrates ranges differ from one content to another.

SD contents are computed from these HD sequences through a half band filtering followed by sub-sampling by a factor of 2 (along horizontal and vertical directions) as shown in Figure 2. This results in nearly SD 576i sequences with a resolution of  $960 \times 540$ . This technique is motivated by the fact that this resolution is very close to “real 16/9” SD ( $1024 \times 576$ ) and that no interpolation is required to convert HD videos to SD videos. With sub-sampling, samples positions do not move. Furthermore, this results in a half-height video (QHD in figure 2, for Quarter HD), which allows to respect both recommended distances for SD ( $D = 6H$ ) and HD ( $D = 3H$ ),  $H$  being the video’s height.

To avoid screen flickering and screen’s manual switching between HD and SD, SD videos have been inserted at the center of an HD video. To do it, gray borders surrounded the SD video (Y level = 73, which corresponds to 200 mV electric video signal as specified in ITU recommendations



**Fig. 3.** Example of SD image inserted in a HD image

BT.500-11 [6] and BT.710-4 [7]). An example of a SD image inserted in a HD image is presented in Figure 3.

As HD contents, SD videos have been encoded with H.264/AVC JM reference software, with the same parameters. We chose not to use MPEG-2 to avoid introducing another difference in the comparison. Each SD sequence has been encoded with two bitrates corresponding to two common SD broadcast qualities. These qualities have been chosen to be representative of an excellent and of a rather good subjective quality respectively. It means to get scores of around 80 and 60 on a continuous subjective quality scale. These will be compared to the 7 different HD levels of quality (from bad to excellent) due to H.264/AVC encoding.

### 3.3. Protocol

We designed specific protocols for these tests since no such experiments have been normalized yet. The protocol is derived from the comparison method with adjectival categorical judgment described in BT.500-11 ITU Recommendation [6]. To determine the observation distance measurement, we ask the observer to move his seat in such a manner that he is comfortably installed to watch HD content. Distance from the seat to the screen is then measured. This measurement should not take more than two or three minutes. Then the observer is asked to take place at three times the height of the screen, as it is the HDTV recommended observation distance. The core of the “HD versus SD” test may then begin.

A test session is made of several presentations. A presentation is made of one or several visualizations of two video sequences labeled “video A” and “video B”. HD and SD videos are assigned letter A or B in random order. A visualization is the viewing of the two videos A and B. During each visualization, the observer compares these two videos. After each presentation, the observer has to report the existence and direction of perceptible differences he perceived. The comparison scale used is shown in Table 1. Values stored and used for the analysis are not shown to the ob-

Caption to choose	Value stored
I prefer much more A than B	+3
I prefer more A than B	+2
I prefer a little more A than B	+1
I have no preference	0
I prefer a little less A than B	-1
I prefer less A than B	-2
I prefer much less A than B	-3

**Table 1.** Comparison scale for “HD vs SD” protocol

server. We choose not to use words with quality connotation like *better* or *worse* as in ITU Recommendations. This way, the user report his global preference, not only with a quality criterion.

For each SD quality level, observers assess every HD qualities. So, the votes analysis will tell how the HD content is preferred over the SD content. The results of these tests will give the required bitrate for each HD content in order to obtain a preference level for HD compared to SD. By this way we can find correspondances between HD and SQ quality scale.

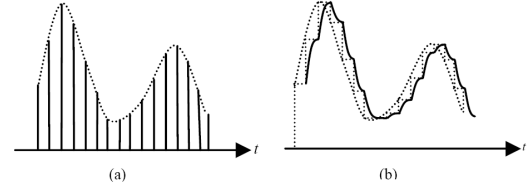
#### 4. MOTION BLUR PERCEPTION ON LCD DISPLAYS

Many LCD defects have been detected by viewers in recent subjective quality tests [3]. Despite recent improvements, motion blur remains still annoying for moving pictures with quick movement. It is particularly noticeable for horizontal movements. According to these considerations, visual motion blur measurement has been considered.

##### 4.1. Motion blur

LCD motion blur is caused jointly by the slow temporal response and by the hold-type LCD’s displaying method. Slow temporal response is due to technology and depends directly on the response time of the crystal from the command. Recent methods like *response time compensation* (RTC) [8] have permitted to reduce considerably the temporal response of LCD matrices. However, even if the response time was null, the blur introduced by motion would not be removed at all [9].

In fact, the most significant cause of the motion blur is the displaying method of LCDs. Emitted light is sustained in the frame period of the video signal like in Figure 4b. LCDs are therefore called hold-type displays. This displaying method is different from CRT where it consists in pulses like in Figure 4a. Then, the perception of motion blur can be explained by two properties of the human visual system (HVS). First is the fact that human eyes are able to track



**Fig. 4.** Temporal waveform of a pixel on CRT (a) and LCD (b), from Pan *et al.* [10].

moving objects perfectly. Secondly, the light stimulus is entirely integrated by the HVS over the frame period time.

During a frame period, moving image is sustained on the screen. Objects in movement stay immobile but the eye continues to move slightly, anticipating the movement. Edges of the object are then integrated on the retina during the whole frame period, resulting in a perceptual blur effect.

To measure the influence of HVS in motion blur, Pan *et al.* [10] have developed a simple mathematical model in which the temporal response of liquid crystal is a parameter. This model is designed to predict the edge perception of a moving object on a LCD device. For a sinusoidal type of response, this model predicts a blur width of  $1.044vT$ , with  $v$  the velocity of the object in pixel per second, and  $T$  the frame period. The blur width is measured between 10% and 90% of signal magnitude. It is defined as the size of the blurred area perceived in the direction of the movement.

Their model also permits to predict the motion blur perceived with an ideal response time of zero. Then, the blur width is about  $0.8vT$ . It can be concluded that HVS motion tracking function associated to a hold-type LCD’s displaying method is responsible for 70 to 80% of the motion blur while slow response time is responsible for only 20 to 30%.

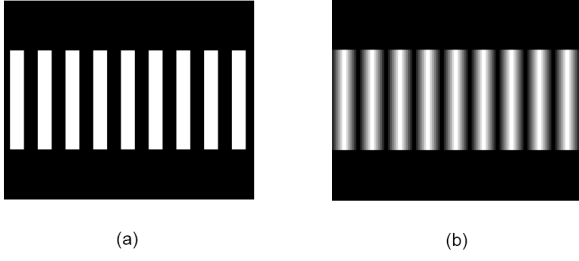
In order to validate Pan’s sinusoidal model and to develop our own model, we have designed and realized subjective tests permitting to measure blur width as a function of motion speed. Experiments and results are described below.

##### 4.2. Subjective experiments

###### 4.2.1. Conception

Experiments have been conducted in order to measure blur width as a function of motion speed. However, the perception of motion blur is directly related to the tracking of the moving object. If the observer stops tracking, to measure blur for example, then the blur is not perceived anymore. That’s why we had to design a test in which the measure of the blur is done while perceiving it.

Experiments consist in displaying a periodical structure of bars moving on a black background at a constant speed.



**Fig. 5.** Displayed (a) and perceived (b) images for a horizontal movement from left to right.

The scrolling is continuous. Due to motion blur, edges of the bars don't appear sharp like shown in Figure 5a but spread in the gap between two bars like in Figure 5b. During the test, the observer has to modify the space between the bars until the two blurred areas begin to blend together. The space between two bars for which two blurs are just merging gives us the width of the motion blur.

#### 4.2.2. Protocol

Viewing conditions are nearly the same as those described in the BT.710-4 ITU-R Recommendation [7] except the part concerning the screen dimensions. Assessment display device is a 17-inch DELL monitor (E172FP) used at native resolution of  $1280 \times 1024$  pixels, with a refresh frequency of 75 Hz. Seven observers took part in these measures. Five of them are 20-year old, two others are 30-year old and 60-year old. All of them were familiar with the procedure and have a perfectly corrected sight.

A session consists in a set of 17 presentations, with four types of stimuli shown in first column of Table 2. The order of visualization of the presentations is random.

For a given presentation, the scrolling of the bars is continuous. Using the arrow keys of a keyboard, the observer can increase or decrease, in real-time, the space between the bars. He can operate as many times as he wants, until he considers that the two blurred areas are just merging. He then validates his measure and the next presentation is displayed. The length of a session varies from an observer to another, but the average time is between 10 and 15 minutes.

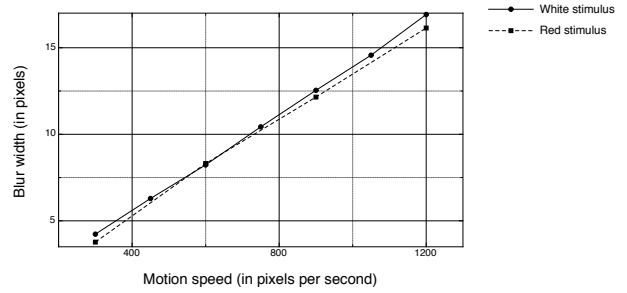
Each of the seven observers have repeated the test twice, on different days. We finally obtain a set of 14 observations for each stimulus.

### 4.3. Results

Results of these tests are shown in Table 2. In the explored range of speed, the width of blur is proportional to motion velocity, as well for horizontal movement (presented in Figure 6), as for vertical movement (in Figure 7). We

Stimulus	Motion speed	Blur width
horizontal white bars	300	4.23
	450	6.29
	600	8.23
	750	10.43
	900	12.54
	1050	14.57
	1200	16.92
horizontal red bars	300	3.77
	600	8.31
	900	12.15
	1200	16.14
vertical white bars	300	4.42
	450	6.50
	600	8.50
vertical red bars	300	4.17
	450	6.33
	600	8.33

**Table 2.** Results of subjective tests : blur width (in pixels) as a function of motion speed in pixels per second. (mean opinion score of seven observers).



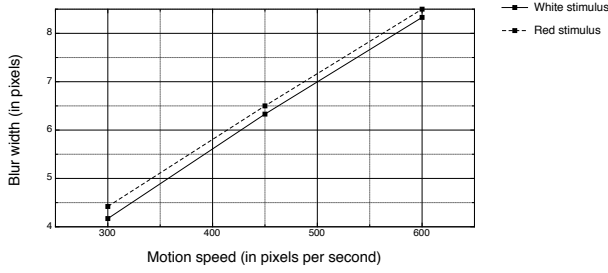
**Fig. 6.** Mean opinion score of observations for horizontal stimulus.

observe that there is no significant differences between the four types of stimuli.

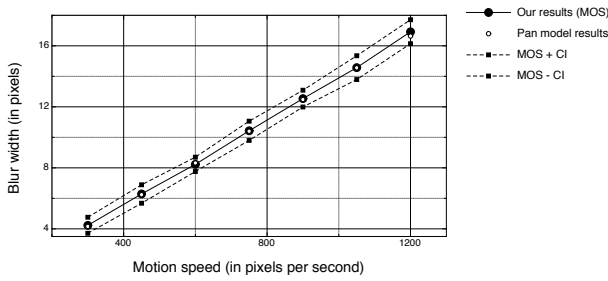
Figure 8 presents the comparison between our results for a horizontal white stimulus and the Pan's sinusoidal model. The correlation coefficient is equal to 0.940. The objective model has a very good correlation with our subjective experiments. We can then conclude that the model is a good approximation of the motion blur perception induced both by hold-type rendering of LCD displays and spatio-temporal human visual system behavior.

## 5. CONCLUSION

In this paper, we presented what HDTV environment implies in terms of quality assessment. New protocols and



**Fig. 7.** Mean opinion score of observations for vertical stimulus.



**Fig. 8.** Comparison between our results for a horizontal white stimulus and Pan's model (MOS is the mean opinion score, and CI is the 95% confidence interval).

metrics are needed to take in account the new features of this broadcasting system. Since no specific work has been done yet on HDTV quality assessment, we presented our own protocols for subjective tests. First, we proposed a protocol to identify the quality gap between HDTV and SDTV. This is an important knowledge for broadcasters to determine the quality improvement due to HD. Then, we presented a protocol to build a motion blur perception model. We designed subjective tests to evaluate blur width as a function of motion speed. Some results from these tests were presented and compared to the ones of a pre-existent model.

## 6. ACKNOWLEDGMENT

This work is supported by HD4U European project. The aim of HD4U is to study best conditions to deploy HDTV in Europe. Several actors from consumer devices industries (Philips, Thomson, etc.) and broadcasters (TF1, Euro1080) are involved. Authors would also like to thank SVT for the open HDTV sequences, Thomson for the HDTV screen and Arnaud Tirel for his assistance in performing the experiments described in the paper.

## 7. REFERENCES

- [1] Ichiro Yuyama, "Fundamental requirements for High-definition television systems – Large-screen effects," NHK technical monograph, NHK, 1982.
- [2] Tetsuo Mitsuhashi, "Fundamental requirements for High-definition television systems – Scanning specifications and picture quality," NHK technical monograph, NHK, 1982.
- [3] ITU, "Report on results of comparative subjective picture quality assessment test between CRT and LCD," Questions ITU-R 95/6, 102/6, ITU – Radiocommunication Study Groups, 2005.
- [4] SVT, "Overall-quality assessment when targeting Wide-XGA flat panel displays," Tech. Rep., SVT corporate development technology, 2002.
- [5] Joint Video Team (JVT), "H.264/Advanced Video Coding reference software version 10.1," 2005, <http://iphone.hhi.de/suehring/tml/>.
- [6] ITU-R BT. 500-11, "Methodology for the subjective assessment of the quality of television pictures," Tech. Rep., International Telecommunication Union, 2004.
- [7] ITU-R BT. 710-4, "Subjective assessment methods for image quality in high-definition television," Tech. Rep., International Telecommunication Union, 1998.
- [8] Richard I. McCartney, "A liquid crystal display response time compensation feature integrated into an lcd panel timing controller," *SID Symposium Digest of Technical Papers*, vol. 34, no. 1, pp. 1350–1353, 2003.
- [9] Taiichiro Kurita, "Moving picture quality improvement for hold-type AM-LCDs," *SID Symposium Digest of Technical Papers*, vol. 32, pp. 986–989, 2001.
- [10] Hao Pan, Xiao-Fan Feng, and Scott Daly, "LCD motion blur modeling and analysis," *IEEE International Conference on Image Processing*, vol. II, pp. 21–24, 2005.